Supplement to "Bayesian Consensus Clustering"

Eric F. Lock[1,2] and David B. Dunson[1]

[1]Department of Statistical Science, Duke University, Durham, NC 27708, U.S.A.
[2]Center for Human Genetics, Duke University Medical Center, Durham, NC 27710, U.S.A.

Abstract

This supplemental article provides additional details and validation for Bayesian consensus clustering (BCC). Section 1 gives full computational details for the algorithm under a normal-gamma model with cluster-specific mean and variance. Section 2 illustrates the relationship between overall and source-specific cluster sizes. Section 3 shows the equivalence of BCC and MDI, under certain assumptions, when $M = 2$ and $K = 2$. Section 4 illustrates the effect of the prior for α. Section 5 provides full details for the clustering comparison simulation in the main article. Section 6 provides computational details and further analysis for the application to TCGA data given in the main article.

1 Computational details

1.1 Normal-gamma model

Here we fill in the details for a specific case of the Bayesian computational framework given in Section 4 of the main article. We assume $\mathbb{X}_i$ has a normal-gamma mixture distribution with cluster-specific mean and variance. That is,

$$X_{mn}|L_{mn} = k \sim N(\mu_{mk}, \Sigma_{mk}),$$

where

- μ_{mk} is a D_m dimensional mean vector, where D_m is the dimension of data source m.
- Σ_{mk} is a $D_m \times D_m$ diagonal covariance matrix, $\Sigma_{mk} = \text{Diag}(\sigma_{mk1}, \ldots, \sigma_{mkD_m})$.

We use a D_m dimensional normal-inverse-gamma prior distribution for $\theta_{mk} = (\mu_{mk}, \Sigma_{mk})$. That is,

$$\theta_{mk} \sim N\Gamma^{-1}(\eta_{m0}, \lambda_0, A_{m0}, B_{m0}),$$

where η_{m0}, λ_0,A_{m0} and B_{m0} are hyperparameters. It follows that μ_{mk} and Σ_{mk} are given by

- $\frac{1}{\sigma^2_{mkd}} \sim \text{Gamma}(A_{m0d}, B_{m0d})$, and
- $\mu_{mkd} \sim N(\eta_{m0}, \frac{\sigma^2_{mkd}}{\lambda_0})$ for $d = 1, \ldots, D_m$.

By default we set $\lambda_0 = 1$, and estimate μ_{m0}, A_{m0} and B_{m0} from the mean and variance of each variable in $\mathbb{X}_m$.

The i'th iteration in the MCMC sampling scheme procedes as follows:

1. Generate $\Theta^{(i)}_m$ given $\{\mathbb{X}_m, \mathbb{L}^{(i-1)}_m\}$, for $m = 1, \ldots, M$. The posterior distribution for $\theta^{(i)}_{mk}$, $k = 1, \ldots, K$ is

   $$\theta^{(i)}_{mk} \sim N\Gamma^{-1}(\eta^{(i)}_{mk}, \lambda^{(i)}_k, A^{(i)}_{m0}, B^{(i)}_{m0}).$$

   Let N_{mk} be the number of samples allocated to cluster k in $\mathbb{L}^{(i-1)}_m$, $\bar{X}_{mk}$ be the sample mean vector for cluster k, and S_{mk} the sample variance vector for cluster k. The posterior normal-gamma parameters are

   - $\eta^{(i)}_{mk} = \frac{\lambda_0 \eta_{m0} + N_{mk}\bar{X}_{mk}}{\lambda_0 + N_{mk}}$
   - $\lambda^{(i)}_k = \lambda_0 + N_{mk}$
   - $A^{(i)}_{m0} = A_{m0} + \frac{n}{2}$
   - $B^{(i)}_{m0} = B_{m0} + \frac{N_{mk}S_{mk}}{2} + \frac{\lambda_0 N_{mk}(\bar{X}_{mk} - \mu_{m0})^2}{2(\lambda_0 + N_{mk})}$.

2. Generate $\mathbb{L}^{(i)}_m$ given $\{\mathbb{X}_m, \Theta^{(i)}_m, \alpha^{(i-1)}_m, \mathbb{C}^{(i-1)}\}$, for $m = 1, \ldots, M$. The posterior probability that $L^{(i)}_{mn} = k$ for $k = 1, \ldots, K$ is proportional to

   $$\nu(k, C^{(i-1)}_n, \alpha^{(i-1)}_m) f_m(X_{mn} | \theta^{(i)}_{mk}),$$

   where f_m is the multivariate normal density defined by $\theta_{mk} = (\mu_{mk}, \Sigma_{mk})$.

3. Generate $\alpha_m^{(i)}$ given $\{\mathbb{C}^{(i-1)}, \mathbb{L}_m^{(i)}\}$, for $m = 1, \ldots, M$. The posterior distribution for $\alpha_m^{(i)}$ is TBeta$(a_m + \tau_m, b_m + N - \tau_m, \frac{1}{K})$, where τ_m is the number of samples n satisfying $L_{mn}^{(i)} = C_n^{(i-1)}$.

4. Generate $\mathbb{C}^{(i)}$ given $\{\mathbb{L}_m^{(i)}, \Pi^{(i-1)}, \alpha^{(i)}\}$. The posterior probability that $C_n^{(i)} = k$ for $k = 1, \ldots, K$ is proportional to

$$\pi_k \prod_{m=1}^{M} \nu(k, L_{mn}^{(i)}, \alpha_m^{(i)}).$$

5. Generate $\Pi^{(i)}$ given $\mathbb{C}^{(i)}$. The posterior distribution for $\Pi^{(i)}$ is Dirichlet $(\beta_0 + \rho)$ where ρ_k is the number of samples allocated to cluster k in $\mathbb{C}^{(i)}$.

By default, we initialize $\mathbb{L}_1, \ldots, \mathbb{L}_k$ by a K-means clustering of each dataset. After running the Markov chain for a specified number of iterations (e.g., 10000), the method described in [2] is used to determine a hard clustering for each of $\mathbb{C}$, $\mathbb{L}_1, \ldots, \mathbb{L}_m$.

For massive datasets parallel computating techniques may be used for the computationally intensive steps (1) and (2) within each MCMC iteration. However, these capabilities are not used in our current implementation.

1.2 Assuming equal adherence

It is straightforward to modify the procedure above under the assumption that each data source adheres equally well to the overall clustering $\mathbb{C}$. Rather than modeling $\alpha_1, \ldots, \alpha_m$ separately, we assume $\alpha = \alpha_1 = \ldots = \alpha_m$. The prior for α is a truncated beta distribution:

$$\alpha \sim \text{TBeta}(a, b, \frac{1}{K}).$$

The MCMC sampling scheme procedes exactly as above, except that in step (3) we need only generate $\alpha^{(i)}$ from the posterior distribution TBeta$(a+\tau, b+NM-\tau, \frac{1}{K})$. Here $\tau = \sum_{m=1}^{M} \tau_m$, where the τ_m are as defined in step (3) above.

2 Cluster size illustration

Recall that π_k is the marginal probability that an object belongs to the overall cluster k:

$$\pi_k = P(C_n = k).$$

The probability that an object belongs to a given source-specific cluster is then

$$P(L_{mn} = k|\Pi) = \pi_k \alpha_m + (1 - \pi_k)\frac{1 - \alpha_m}{K - 1}.$$

As a consequence, the size of the source-specific clusters are generally more uniform than the size of the overall clusters. In particular, the source-specific clusterings $\mathbb{L}_m$ will generally represent more clusters than $\mathbb{C}$, rather than vice-versa.

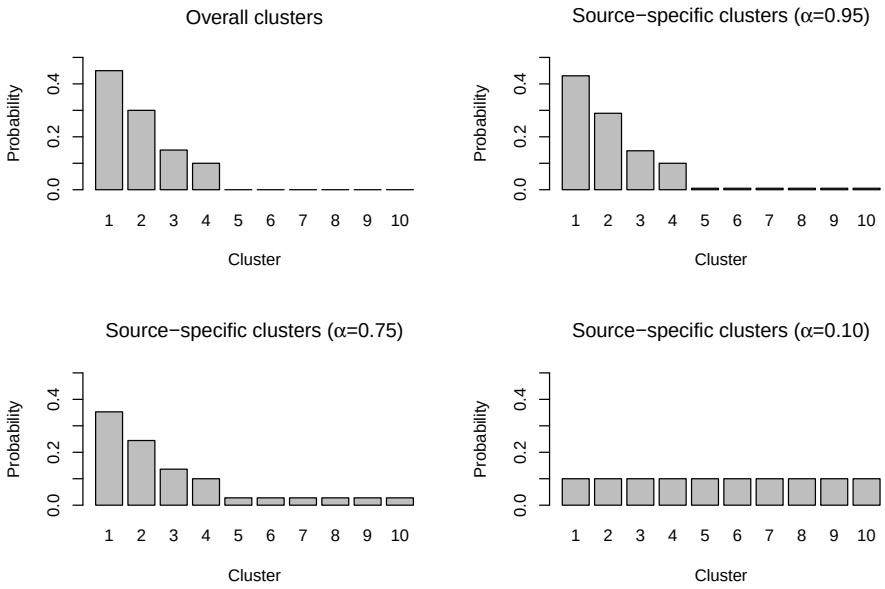

Figure 1: Marginal cluster inclusion probabilities are shown for the overall clusters (Π) with $K = 10$ (top-left). The source-specific cluster probabilities induced by Π are shown for the adherence levels $\alpha_m = 0.95, 0.75$ and 0.10.

As an illustration, we set $K = 10$ and assume the τ_k have the skewed distribution shown in the top left panel of Figure 1. The marginal cluster inclusion probabilities for a given data source depend on its adherence α_m to the overall clustering. Not surprisingly, for α_m close to 1 the inclusion probabilities closely resemble those for the overalll clustering (top right panel of Figure 1). As α_m approaches $\frac{1}{K}$, the inclusion probabilities are more uniform (bottom two panels of Figure 1). In particular, clusters that had zero probability to occur in $\mathbb{C}$ have positive probability to occur in $\mathbb{L}_m$. Hence, a sample that does not fit any overall pattern in a given data source (e.g., an outlier) need not be allocated to an overall cluster.

3 Equivalence of BCC and MDI

Here we compare the MDI and BCC models for $M = 2$ data sources and $K = 2$ clusters, giving conditions where the two models are equivalent under a parameter substitution.

Assume $\alpha = \alpha_1 = \alpha_2$, and let $U = \frac{\alpha}{1-\alpha}$. The joint distribution of $(\mathbb{L}_1, \mathbb{L}_2)$ under BCC is then

$$P(\{L_{mn} = k_m\}_{m=1}^2 | \Pi, \alpha) \propto \begin{cases} \pi_1 U^2 + (1-\pi_1) & \text{if } k_1 = k_2 = 1 \\ \pi_1 + (1-\pi_1)U^2 & \text{if } k_1 = k_2 = 2 \\ U & \text{if } k_1 \neq k_2. \end{cases}$$

Assume $\tilde{\pi} = \tilde{\pi}_{1\cdot} = \tilde{\pi}_{2\cdot}$ and let $\phi = \phi_{12}$ in the MDI clustering model. The joint distribution of $(\mathbb{L}_1, \mathbb{L}_2)$ under MDI is then

$$P(\{L_{mn} = k_m\}_{m=1}^2 | \tilde{\Pi}, \phi) \propto \begin{cases} \tilde{\pi}_1^2(1+\phi) & \text{if } k_1 = k_2 = 1 \\ (1-\tilde{\pi}_1)^2(1+\phi) & \text{if } k_1 = k_2 = 2 \\ \tilde{\pi}_1(1-\tilde{\pi}_1) & \text{if } k_1 \neq k_2. \end{cases}$$

It is straightforward to verify that the two forms are equivalent under the substitutions

$$\phi = \sqrt{\left((1-\pi_1)U + \frac{\pi_1}{U}\right)\left(\pi_1 U + \frac{1-\pi_1}{U}\right)} - 1$$

and

$$\tilde{\pi}_1 = \frac{\sqrt{(1-\pi_1)U^{-1} + \pi_1 U}}{\sqrt{(1-\pi_1)U + \pi_1 U^{-1}} + \sqrt{(1-\pi_1)U^{-1} + \pi_1 U}}.$$

There is no such equivalence for $M > 2$ or $K > 2$, regardless of restrictions on Π and Φ.

4 Prior comparison for α

We use a simple simulation to illustrate the effect of the prior distribution for α. We generate datasets $\mathbb{X}_1 : 1 \times 200$ and $\mathbb{X}_2 : 1 \times 200$ as in Section 5.1 of the main article:

1. Let $\mathbb{C}$ define two clusters, where $C_n = 1$ for $n \in \{1, \ldots, 100\}$ and $C_n = 2$ for $n \in \{101, \ldots, 200\}$

2. Draw α from a Uniform$(0.5, 1)$ distribution.

3. For $m = 1, 2$ and $n = 1, \ldots, 200$, generate $L_{mn} \in \{1, 2\}$ so that $P(L_{mn} = C_n) = \alpha$ and $P(L_{mn} \neq C_n) = 1 - \alpha$.

4. For $m = 1, 2$ draw values X_{mn} from a Normal$(1.5, 1)$ distribution if $L_{mn} = 1$ and from a Normal$(-1.5, 1)$ distribution if $L_{mn} = 2$

We estimate the BCC model under the assumption that $\alpha = \alpha_1 = \alpha_2$, where α has prior distribution TBeta$(a, b, \frac{1}{2})$, for various values of a and b. The uniform prior ($a = b = 1$) gives relatively unbiased results, as illustrated in Figure 1 of the main manuscript. Figure 2 displays the estimated values $\hat{\alpha}$ for alternative choices of a and b. Not surprisingly, for very precise priors (large a and b) the $\hat{\alpha}$ are highly influenced by the prior and are therefore inaccurate. However, the $\hat{\alpha}$ appear to be robust for moderately precise priors.

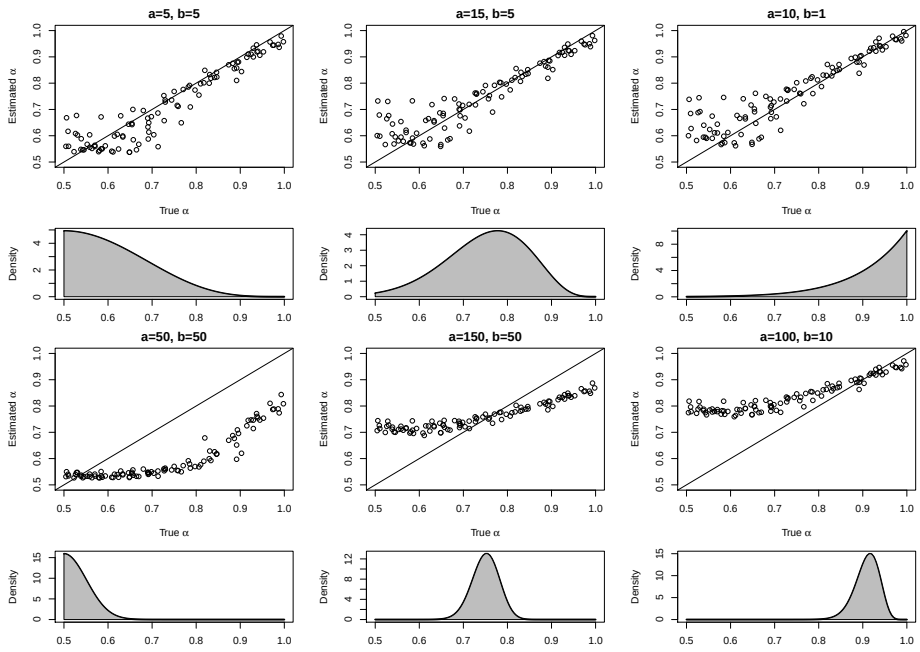

Figure 2: Scatterplots of $\hat{\alpha}$ versus the true α are shown for various prior distributions on α. Each prior is of the form TBeta$(a, b, \frac{1}{2})$, and the density of each prior is shown below the relevant scatterplot.

5 Clustering comparison details

Here we describe the computational details for the four procedures used in the clustering comparison study given in Section 5.2 of the main manuscript.

For each procedure the MCMC algorithm ran for 1000 iterations (after 200 iterations of "burn-in"), and a hard clustering was determined as in Dahl [2].

5.1 Separate clustering

We use a normal-gamma mixture model to cluster each $\mathbb{X}_m$. The marginal probability that X_{mn} is allocated to cluster k is π_{mk}, where $\Pi_m = (\pi_{m1}, \ldots, \pi_{mk}) \sim$ Dirichlet(β_0). We use $K = 2$ clusters and $\beta_0 = (1, 1)$. The i'th iteration in the MCMC sampling scheme procedes as follows:

1. Generate $\Theta_m^{(i)}$ given $\{\mathbb{X}_m, \mathbb{L}_m^{(i-1)}\}$. The posterior distribution for $\theta_{mk}^{(i)}$, $k = 1, \ldots, K$ is

$$\theta_{mk}^{(i)} \sim N\Gamma^{-1}(\eta_{mk}^{(i)}, \lambda_k^{(i)}, A_{m0}^{(i)}, B_{m0}^{(i)}).$$

   Let N_{mk} be the number of samples allocated to cluster k in $\mathbb{L}_m^{(i-1)}$, $\bar{X}_{mk}$ be the sample mean vector for cluster k, and S_{mk} the sample variance vector for cluster k. The posterior normal-gamma parameters are

   - $\eta_{mk}^{(i)} = \frac{\lambda_0 \eta_{m0} + N_{mk}\bar{X}_{mk}}{\lambda_0 + N_{mk}}$
   - $\lambda_k^{(i)} = \lambda_0 + N_{mk}$
   - $A_{m0}^{(i)} = A_{m0} + \frac{n}{2}$
   - $B_{m0}^{(i)} = B_{m0} + \frac{N_{mk}S_{mk}}{2} + \frac{\lambda_0 N_{mk}(\bar{X}_{mk} - \mu_{m0})^2}{2(\lambda_0 + N_{mk})}$.

2. Generate $\mathbb{L}_m^{(i)}$ given $\{\mathbb{X}_m, \Theta_m^{(i)}, \Pi_m^{(i-1)}\}$. The posterior probability that $L_{mn}^{(i)} = k$ for $k = 1, \ldots, K$ is proportional to

$$\pi_{mk} f_m(X_{mn} | \theta_{mk}^{(i)}),$$

   where f_m is the multivariate normal density defined by $\theta_{mk} = (\mu_{mk}, \Sigma_{mk})$.

3. Generate $\Pi_m^{(i)}$ given $\mathbb{L}_m^{(i)}$. The posterior distribution for $\Pi_m i =^{(i)}$ is Dirichlet $(\beta_0 + \rho_m)$ where ρ_{mk} is the number of samples allocated to cluster k in $\mathbb{L}_m^{(i)}$.

5.2 Joint clustering

We use a normal-gamma mixture model to cluster the concatenated dataset

$$\mathbb{X} = \begin{bmatrix} \mathbb{X}_1 \\ \vdots \\ \mathbb{X}_M \end{bmatrix}.$$

The computational details are exactly as in Section 5.1, except that we perform the algorithm on the joint data $\mathbb{X}$ rather than separately for each $\mathbb{X}_m$.

5.3 Dependent clustering

For our dependent clustering model we let $\alpha_{m_1m_2}$ be the probability that $L_{m_1n} = L_{m_2n}$, where $\alpha_{m_1m_2} \sim \text{TBeta}(a, b, \frac{1}{K})$. Hence, we model the clustering dependence between each pair of datasets, rather than adherence to an overall clustering. The marginal probability that X_{mn} is allocated to cluster k is π_{mk}, where $\Pi_m = (\pi_{m1}, \ldots, \pi_{mk}) \sim \text{Dirichlet}(\beta_0)$. We use $K = 2$ clusters,$\beta_0 = (1, 1)$, and $a = b = 1$. The i'th iteration in the MCMC sampling scheme procedes as follows:

1. Generate $\Theta_1^{(i)}$ given $\{\mathbb{X}_m, \mathbb{L}_m^{(i-1)}\}$, as in step (1) of Section 1.1.

2. Generate $\mathbb{L}_1^{(i)}, \ldots, \mathbb{L}_M^{(i)}$ given $\{\mathbb{X}, \alpha^{(i-1)}, \Theta^{(i)}, \Pi^{(i-1)}\}$. For $n = 1, \ldots, N$, the joint posterior probability for $\{L_{mn} = k_m\}_{m=1}^M$ is proportional to

$$\prod_{m=1}^{M} \pi_{mk}^{(i-1)} f_m(X_{mn}|\theta_{mk}^{(i)}) \prod_{m'=m+1}^{M} \nu(k_m, k_{m'}, \alpha_{mm'}^{(i-1)}).$$

3. Generate $\alpha_{m_1m_2}^{(i)}$ given $\{L_{m_1}^{(i)}, L_{m_2}^{(i)}\}$, for all pairs $\{(m_1, m_2) : m_1 = 1, \ldots, M \text{ and } m_2 = (m_1 + 1), \ldots, M\}$. The posterior distribution for $\alpha_{m_1m_2}^{(i)}$ is $\text{TBeta}(a + \tau_{m_1m_2}, b + N - \tau_{m_1m_2}, \frac{1}{K})$, where $\tau_{m_1m_2}$ is the number of samples n satisfying $L_{m_1n}^{(i)} = L_{m_2n}^{(i)}$.

4. Generate $\Pi_m^{(i)}$ given $\mathbb{L}_m^{(i)}$. The posterior distribution for $\Pi_m^{(i)}$ is Dirichlet $(\beta_0 + \rho_m)$ where ρ_{mk} is the number of samples allocated to cluster k in $\mathbb{L}_m^{(i)}$.

5.4 BCC

We implement BCC as described in Section 1, assuming equal adherence. We use $K = 2$ clusters, $\beta_0 = (1, 1)$, and $a = b = 1$.

6 TCGA application details

Here we provide more details for the application to heterogenous genomic data from TCGA that is presented in Section 6 of the main article. We focus on the application of BCC to GE, ME, miRNA and RPPA data for

348 breast cancer tumor samples. For more information on the origin of these data and pre-processing details see [1]. See http://people.duke.edu/~el113/software.html for R code [3] to download and process these data and to completely reproduce the analysis. This file also includes instructions on how to download these data from the TCGA Data Portal.

6.1 Choice of K

We use the heuristic method described in Section 4.1 of the main article to choose the number of clusters K. The full BCC model is estimated as in Section 1, for the potential values $K = 2, \ldots, 10$. We model the α_m separately, set $a_m = b_m = 1$ for each m, and set $\beta_0 = (1, \ldots, 1)$. We run the MCMC sampling scheme for $10,000$ iterations for each K (the first 2000 iterations are used as a "burn-in"). We compute the mean adjusted adherence $\bar{\alpha}_K^*$ for each K. Figure 3 shows a point estimate for $\bar{\alpha}_K^*$ (an average over the MCMC iterations), as well as a 95% credible interval based on the MCMC iterations, as a function of K. The maximum value is obtained for $K = 3$ ($\bar{\alpha}_K^* = 0.57$), although the adherence level is comparable for $K = 4$ and $K = 6$.

6.2 MCMC mixing

We consider the $10,000$ MCMC draws for $K = 3$. Posterior estimation takes ~ 30 minutes on a 2.30 GHz laptop with 4 GB ram. The draws appear to mix well and they converge quickly to a stationary posterior distribution. Figure 4 shows the draws for α_m, $m = 1, \ldots, 4$. These are the estimated adherence to the overall clustering for GE, ME, miRNA and RPPA. They appear to converge within the first 1000 iterations to an approximately stationary distribution. The average values are $\alpha = 0.91$ for GE, $\alpha = 0.69$ for ME, $\alpha = 0.56$ for miRNA and $\alpha = 0.70$ for RPPA. Figure 5 shows the marginal overall cluster inclusion probabilities π_k over the MCMC draws. These also quickly converge to a stationary distribution. The average values are $\hat{\pi}_1 = 0.24$, $\hat{\pi}_2 = 0.28$ and $\hat{\pi}_3 = 0.48$.

6.3 Normality diagnostics

Here we check the assumption that clusters are normally distributed for each data source. To assess the overall distribution within each data source, we standardize values for each feature:

$$z_{mnd} = \frac{X_{mnd} - \bar{X}_{m\cdot d}}{s_{m\cdot d}}$$

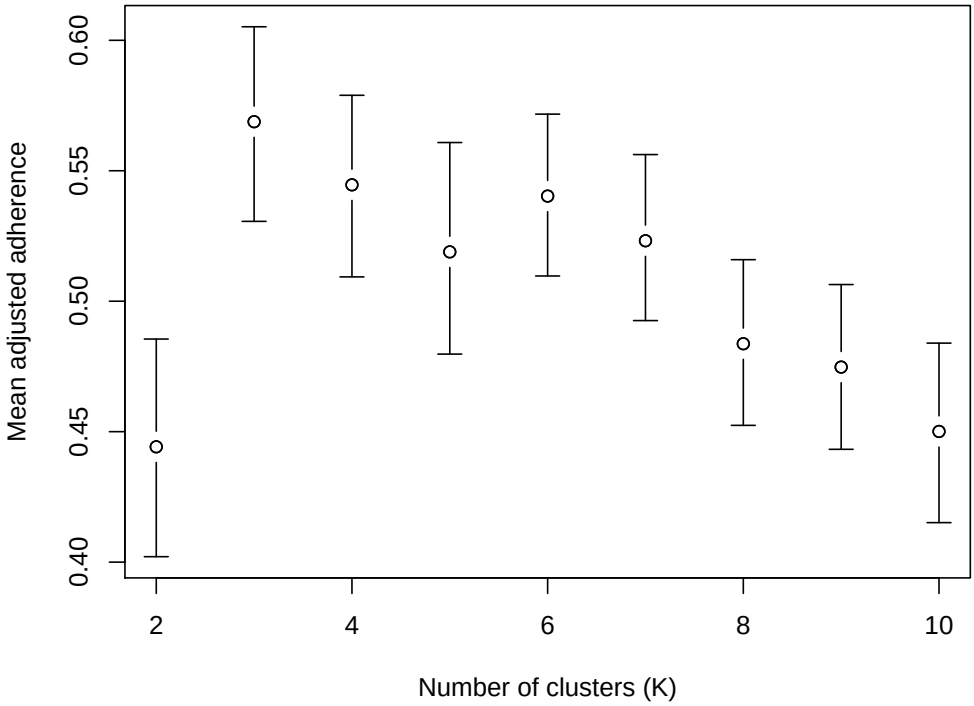

Figure 3: The mean adjusted adherence $\bar{\alpha}_K^*$ is shown after estimating the model with $K = 2, \ldots, 10$. A point estimate given by the average of the MCMC draws, and a credible interval given by the 2.5 and 97.5 percentiles of the MCMC draws, are shown for each K.

where $\bar{X}_{m \cdot d}$ and $s_{m \cdot d}$ are the mean and standard deviation for feature d in data source m. Figure 6 (left) plots the quantiles of the standard normal distribution against the overall quantiles for the standardized values of each data source. These show significant skewness and bimodality. We then standardize data values within each source-specific cluster:

$$z_{mnd} = \frac{X_{mnd} - u_{mL_{mn}d}}{\sigma_{mL_{mn}d}},$$

where μ_{mkd} and σ_{mkd} are the estimated mean and standard deviation for cluster k for feature d in data source m. The quantiles of the cluster-standarized values are shown in Figure 6 (right) and agree closely with the quantiles of the standard normal distribution. The Kolmogorov-Smirnov statistic between

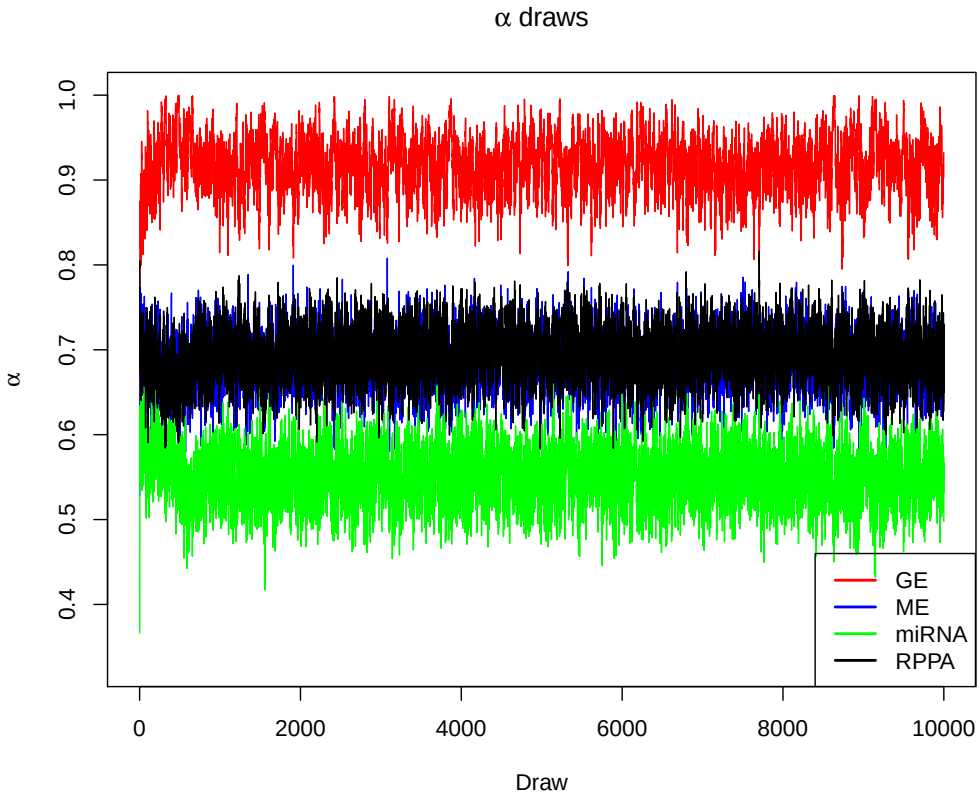

Figure 4: Values of α_m are shown for the data sources GE, ME, miRNA and RPPA over the 10,000 MCMC draws.

the empirical distribution function of the cluster standardized values and the normal distribution was less than 0.05 for each data source ($D = 0.043$ for GE,$D = 0.034$ for ME, 0.022 for miRNA and 0.021 for RPPA). Therefore, the normal distribution seems to be a reasonable cluster model for each data source.

6.4 Comparison with TCGA subtypes

Table 3 of the main article compares the overall clusters identified by BCC with the four overall subtypes defined by TCGA. TCGA has also defined subtypes that are particular to each data source, and we compare the source-specific clusterings identified by BCC with the source-specific subtypes defined by TCGA. Tables 1, 2, 3, and 4 show the cluster vs. subtype matching matrix for GE, ME, miRNA and RPPA, respectively. The subtypes for GE

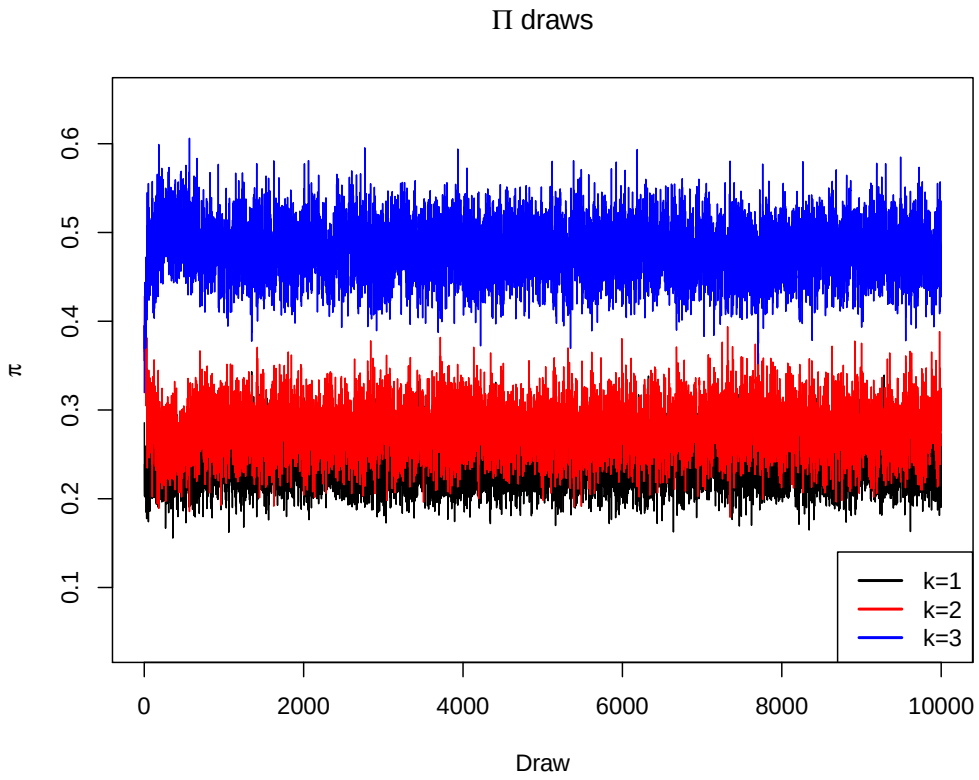

Figure 5: Values of π_k are shown for $k = 1, 2, 3$ over the 10,000 MCMC draws.

and RPPA are named by TCGA according to their biological profile. In all cases the two sample partitions have a significant association (p-value < 0.01; Fisher's exact test). However, these associations are not exceptionally strong, suggesting that the two partitions may not be driven by the same structure in the data.

6.5 Heatmaps

Figures 7, 8, 9, and 10 show heatmaps of the GE, ME, miRNA, and RPPA datasets, respectively. The processed data that was used for clustering is shown in each case. Samples (columns) are grouped by their relevant source-specific cluster. For each heatmap a cluster effect is apparent in a large number of variables (rows).

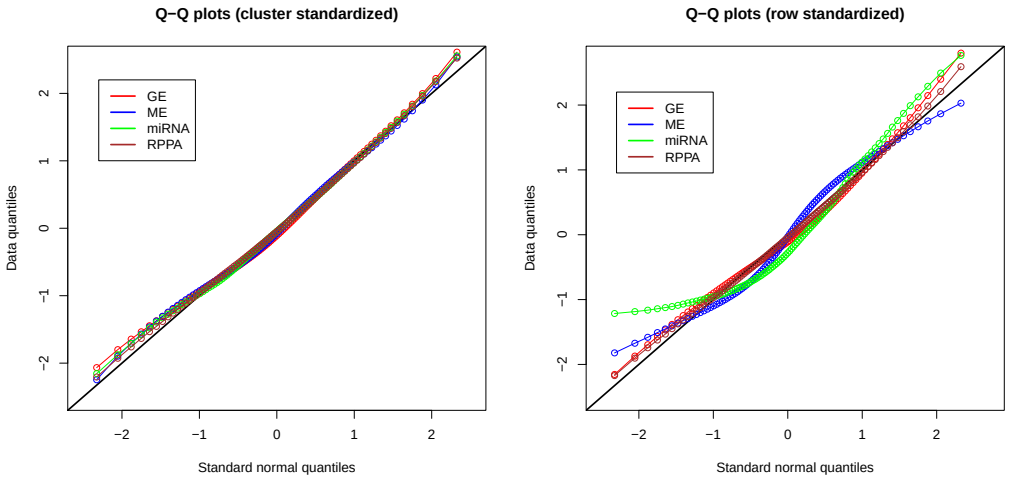

Figure 6: Standard normal quantile-quantile (QQ) plots for each data source, after standardizing values within each row/variable (right) and within each source-specific cluster (left).

		BCC cluster		
GE		**1**	**2**	**3**
	Basal	65	1	0
	HER2	14	5	23
TCGA subtype	**Luminal A**	0	78	76
	Luminal B	0	5	76
	Normal	2	3	0

Table 1: BCC cluster vs. TCGA subtype matching matrix for GE data. The GE subtypes are named according to their biological profile.

		BCC cluster		
ME		**1**	**2**	**3**
	1	0	46	12
	2	1	88	1
TCGA subtype	**3**	0	0	36
	4	0	3	87
	5	73	0	1

Table 2: BCC cluster vs. TCGA subtype matching matrix for ME data.

		BCC cluster		
miRNA		**1**	**2**	**3**
	1	10	9	8
	2	10	14	25
	3	23	4	4
TCGA subtype	**4**	41	66	9
	5	47	3	0
	6	8	43	5
	7	1	8	10

Table 3: BCC cluster vs. TCGA subtype matching matrix for miRNA data.

		BCC cluster		
RPPA		**1**	**2**	**3**
	Basal	61	1	7
	Her2	29	1	13
TCGA subtype	**Luminal A/B**	1	14	105
	ReacI	0	61	2
	ReacII	7	34	0

Table 4: BCC cluster vs. TCGA subtype matching matrix for RPPA data. The RPPA subtypes are named according to their biological profile.

6.6 Survival

Analysis of long-term survival for these data is limited by the relatively short follow-up time (median 1.9 years) and low number of survival events (39 of 348). Nevertheless, both the three clusters identified by BCC and the four comprehensive subtypes identified by TCGA show a difference in short-term survival (Log-rank p-value = 0.04 and 0.001, respectively). Figure 11 gives Kaplan-Meier survival curves for each BCC cluster and TCGA subtype. Among the BCC clusters Cluster 2 has the best prognosis (5-year survival = 0.94), followed by Cluster 3 (0.81) and Cluster 3 (0.67).

References

[1] Cancer Genome Atlas Network (2012). Comprehensive molecular portraits of human breast tumours. *Nature*, **490**(7418), 61–70.

[2] Dahl, D. (2006). *Model-Based Clustering for Expression Data via a Dirichlet Process Mixture Model*. Cambridge University Press.

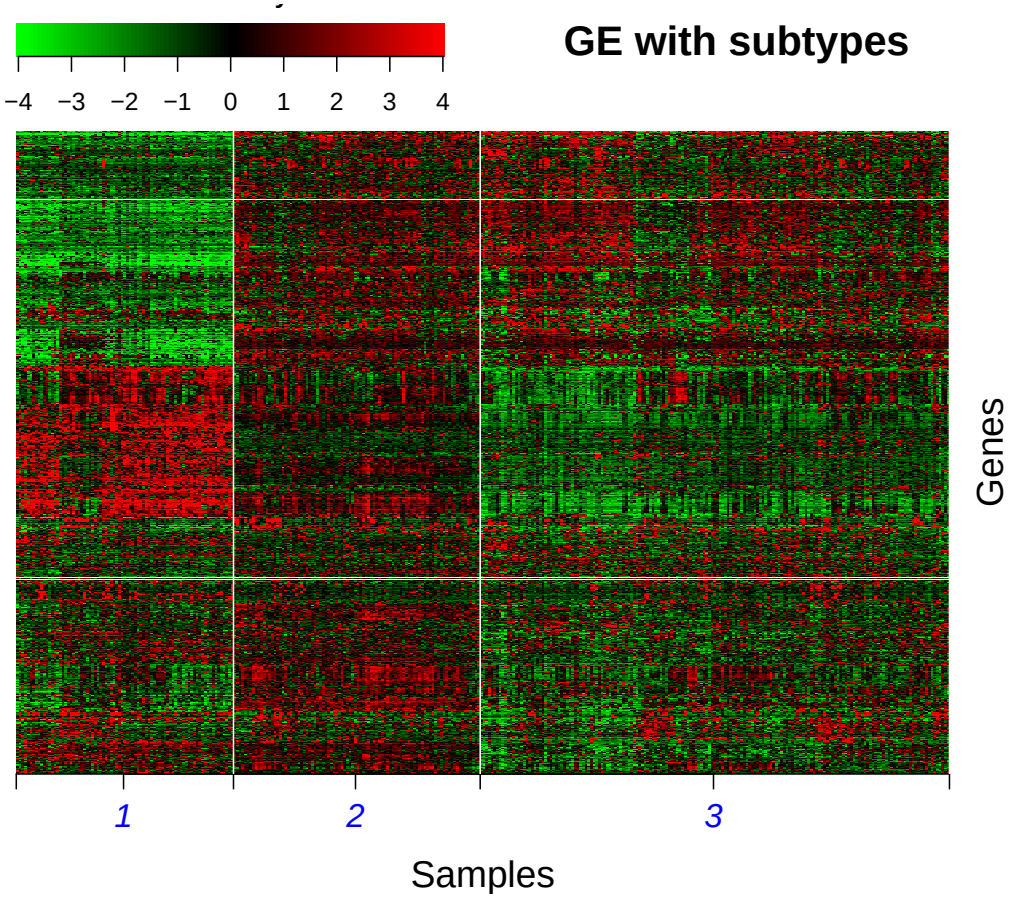

Figure 7: Heatmap of the GE data; samples are grouped by their GE-specific cluster.

[3] R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

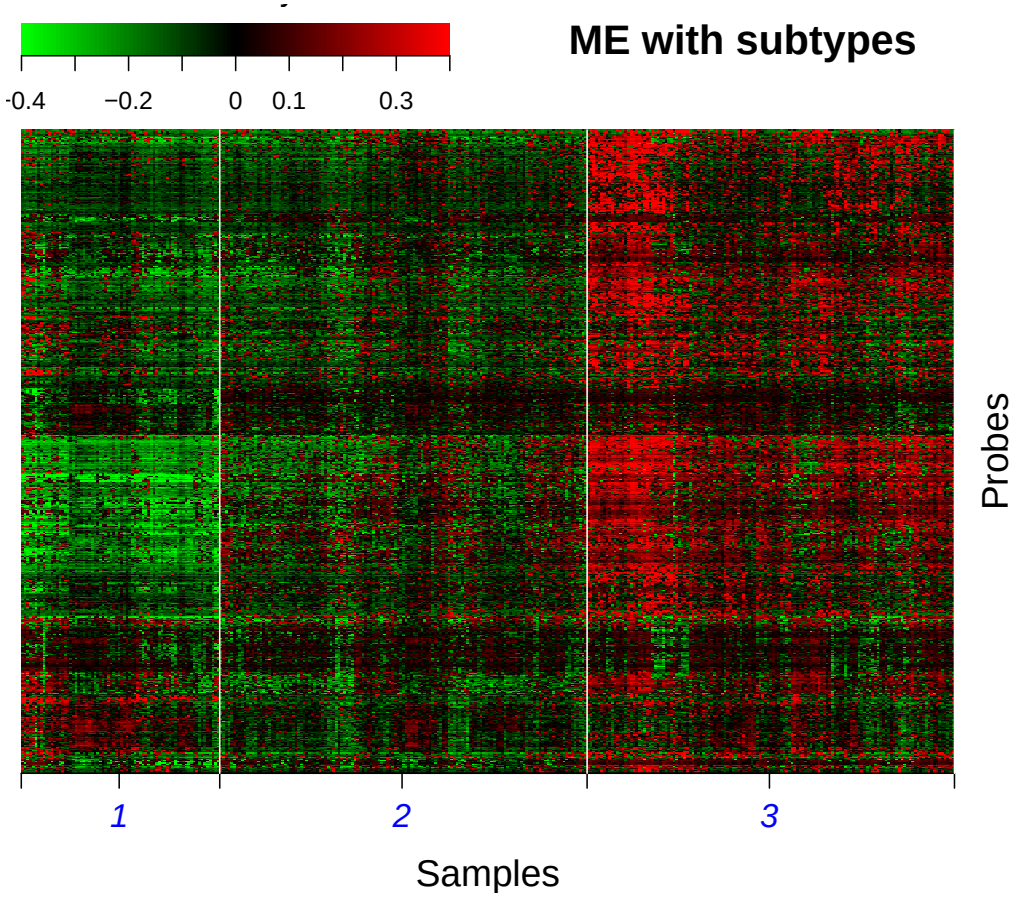

Figure 8: Heatmap of the ME data; samples are grouped by their ME-specific cluster.

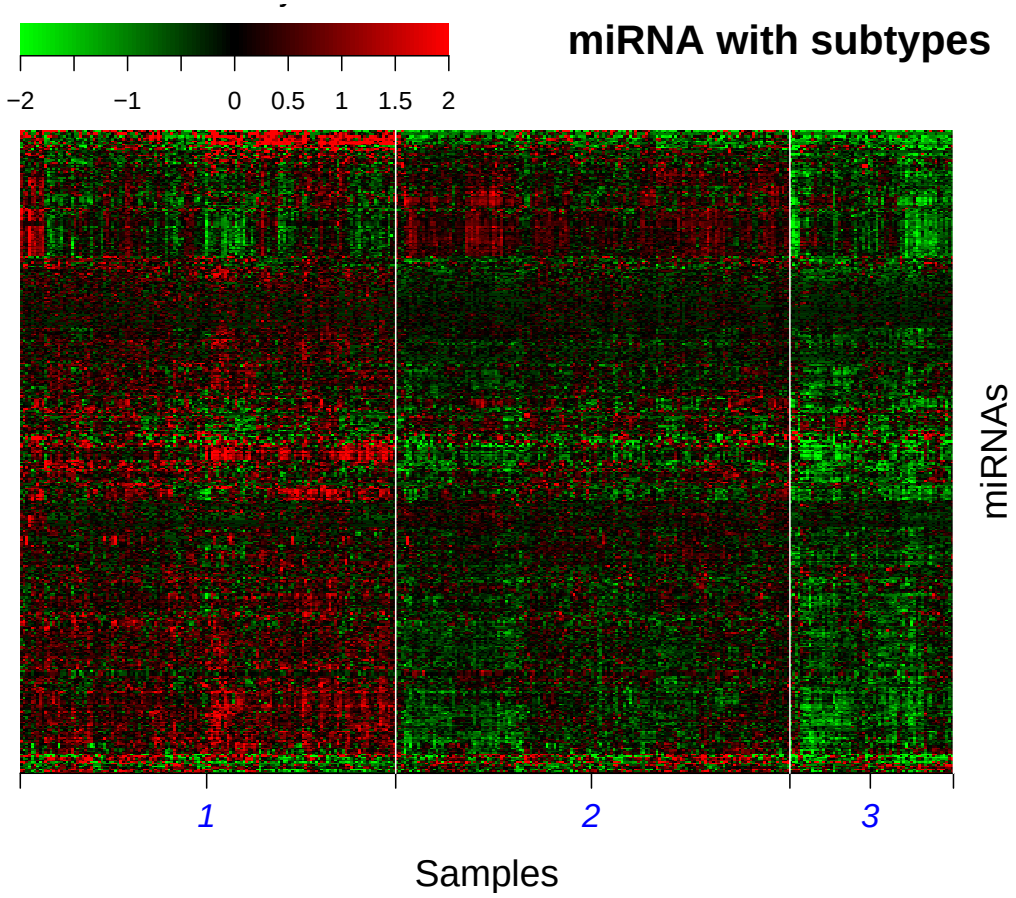

Figure 9: Heatmap of the miRNA data; samples are grouped by their miRNA-specific cluster.

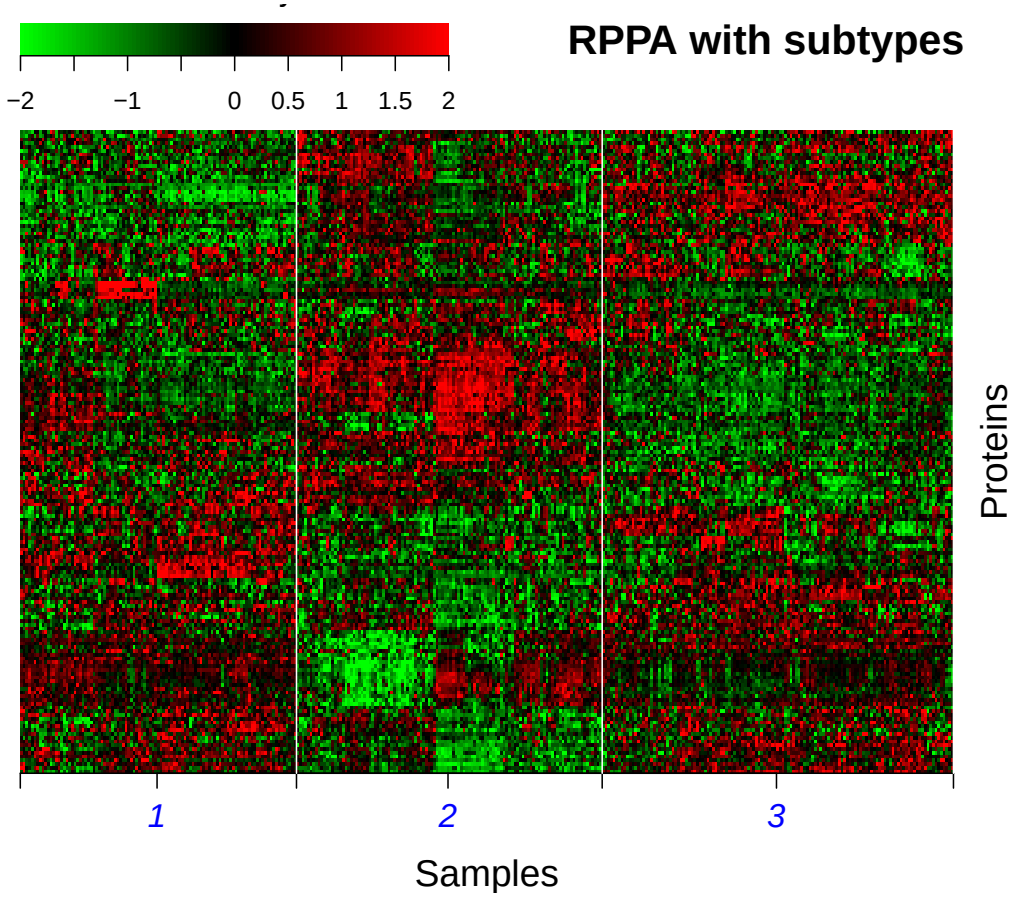

Figure 10: Heatmap of the RPPA data; samples are grouped by their RPPA-specific cluster.

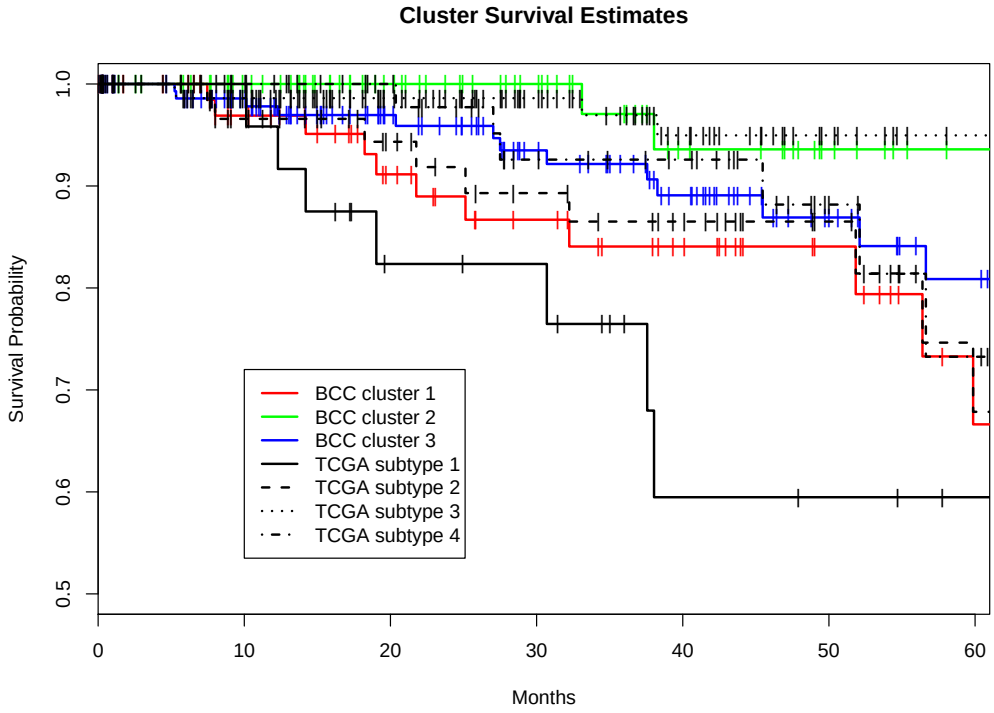

Figure 11: Short-term survival curves for the BCC clusters and TCGA subtypes, using the Kaplan-Meier estimator.